DOCUMENT RESUME

ED 457 746 HE 034 399

AUTHOR DePaolo, Concetta A.
TITLE An Implementation of a Mathematical Programming Approach to Optimal Enrollments. AIR 2001 Annual Forum Paper.
PUB DATE 2001-06-00
NOTE 32p.; Paper presented at the Annual Meeting of the Association for Institutional Research (41st, Long Beach, CA, June 3-6, 2001).
PUB TYPE Reports - Evaluative (142) -- Speeches/Meeting Papers (150)
EDRS PRICE MF01/PC02 Plus Postage.
DESCRIPTORS Decision Making; *Educational Planning; *Enrollment; Higher Education; Mathematical Models
IDENTIFIERS *Mathematical Programming; *Optimization

ABSTRACT

This paper explores the application of a mathematical optimization model to the problem of optimal enrollments. The general model, which can be applied to any institution, seeks to enroll the "best" class of students (as defined by the institution) subject to constraints imposed on the institution (e.g., capacity, quality). Topics explored include how the model was applied to actual data and the results of that application. The presentation touches on how well the model mimics "real life," insights that can be gained from the output, the model's limitations, and the modifications that might be warranted to improve performance. (Contains 17 references.) (Author/SLD)

Reproductions supplied by EDRS are the best that can be made
from the original document.

Running Head: Implementation of a Mathematical Programming Approach

ED 457 746

An Implementation of a Mathematical Programming Approach To Optimal Enrollments

Concetta A. DePaolo

Indiana State University

U.S. DEPARTMENT OF EDUCATION
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
This document has been reproduced as received from the person or organization originating it.
Minor changes have been made to improve reproduction quality.
Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

BEST COPY AVAILABLE

An Implementation of a Mathematical Programming Approach To Optimal Enrollments

Abstract

This paper explores the application of a mathematical optimization model to the problem of optimal enrollments. The general model, which can be applied to any institution, seeks to enroll the "best" class of students (as defined by the institution) subject to constraints imposed upon the institution (e.g. capacity, quality). Topics explored include how the model was applied to actual data and the results of that application. The presentation will touch upon how well the model mimics "real life," insights that can be gained from the output, the model's limitations, and what modifications may be warranted to improve performance.

An Implementation of a Mathematical Programming Approach To Optimal Enrollments

Enrollment management and enrollment projections are important concepts that have been studied extensively in the literature. Statistical and other quantitative techniques for enrollment projections have long been employed (see Brinkman & McIntyre, 1997; Bruggink & Gambhir, 1996; Caruthers & Wentworth, 1997; Conner, 1971; Donhardt, 1995; Schellenberg, & Stephens, 1987). However, while some authors such as Conner (1971), Donhardt (1995), and Brinkman (1997) mention the effects of admissions policies on enrollment, none deal explicitly with implementing those policies.

On the other hand, the topic of optimal enrollment is given much attention in enrollment management settings, however, most studies approach the problem qualitatively (for example, Hossler, Bean, & Associates, 1990; Ingersoll, 1988; Inhlanfeldt, 1980; Day, 1997; Kemerer, Baldridge, & Green, 1982). Due to the complexity of the problem, a qualitative decision-making process can be overwhelming. A quantitative approach to this problem has the benefit of being able to handle many variables and constraints at once, and to uncover patterns that may be difficult to discern using qualitative approaches.

In a study by Averill and Suttle (1975), the authors introduce the idea of a mathematical programming approach to produce optimal enrollment levels. A model examined by DePaolo (2000) uses similar mathematical optimization techniques. This model assumes that a decision must be made each year as to how many students of various types should be admitted, assuming that students of the same type will act similarly with respect to enrollment, persistence and graduation. The mathematical model of the problem takes into account the uncertainty involved; that is, that students who are admitted may or may not enroll, persist, or graduate. This approach

also allows the institution to "optimize" admissions decisions based on various objectives and goals, while satisfying certain constraints imposed on the institution.

This paper will refine the mathematical model introduced by DePaolo by describing additional objectives and goals and simplifying some of the assumptions made in the original formulation. Then, the application of this model to actual data will be discussed and analyzed to illustrate some properties of the "optimal" solution to such a problem. Specifically, it can be shown that for each institutional objective, there is an ordering of student types from "best" to "worst." Finally, a simple way to implement the model using Excel will be described.

The Original Model

The original model introduced is based on the assumption that the uncertainty associated with the behavior of future students is similar to the behavior of past students of the same "type." Student types can be based on academic qualifications, enrollment in various programs, in state vs. out-of-state residence, on campus vs. off campus residence, or any other factors that an institution might consider when making admissions decisions. Another assumption of the original model is that the behavior of students of a given type, for example in yield, retention and graduation rates, does not vary significantly from year to year. It will be discussed in the next section how this assumption may be relaxed.

Based on these assumptions, the model assumes that an institution has various objectives or goals that it would like to achieve, subject to constraints imposed upon the institution. These objectives offer a way to compare one group of students to another in order to determine which is more desirable. The possible objectives put forth in the original model include: (1) quality of the freshman class, measured in terms of pre-entry attributes or post-enrollment behavior, (2) revenues generated by the group of students over their entire tenure at the institution, (3) a combination of quality and revenues, (4) a measure of how well a group of students conforms to

some ideal class, (5) a measure of how well a group of students meets various institutional goals, and (6) a measure of desirability of the class with respect to some general utility function that captures intangible characteristics of students (for example, students might participate in campus activities and contribute to the institution in a way that cannot easily be measured). Although these objectives were all discussed in the initial model, for simplicity, the only one that was discussed extensively was revenue generated. Note that these measures of desirability all depend on uncertain quantities, namely which students will enroll if admitted. We discuss later how this uncertainty is handled in the model.

While an institution has certain objectives or goals it would like to meet, it is also burdened with various constraints. For example, the institution cannot admit more students of a given type than it has applications of a given type. These are termed "real" constraints, as they are a result of the physical attributes of the problem. There are other constraints that have been termed "capacity" constraints because they deal with how many students overall or of a given type the institution can realistically handle. For example, an institution may not be able to handle a freshman class larger than a certain upper limit, or a certain special academic program may not be able to be maintained with fewer than a certain number of enrollees.

The remaining constraints discussed in the original model are called "quality" constraints because they deal with the retention and graduation rates of the incoming students. For example, an institution may want to ensure that its retention and graduation rates are acceptably high before deciding to admit a group of students. Note that the capacity and quality constraints are different from the real constraints because they involve uncertain quantities representing how many students will enroll. Thus, these constraints cannot be met with 100% certainty, but the model allows for this limit to be met with a high degree of probability, for example, 95%.

Since the number of students enrolling is an uncertain quantity, the model seeks to describe this quantity in terms of probability distributions. If we assume each student acts independently of each other, and if admitted, that each student enrolls with probability y_i (the yield rate of student type i), then the number of students of type i who enroll in their first year has a Binomial distribution with $n = x_i$ (the number of students of type i who are admitted) and $p = y_i$. Note that a Binomial distribution is characterized by n trials of an experiment, each of which can result in success (enrollment) with probability p or failure with probability $(1-p)$. Thus, we can similarly describe the number of students enrolling in their $(k+1)^{st}$ year as a Binomial distribution with $n = x_i$ as before, and $p = y_i R_i^k$, where R_i^k is the k^{th} -year retention rate for students of type i.

If we assume that each student type contains a sufficient number of students, then such a Binomial distribution can be approximated with a normal distribution with mean $\mu = np$ and variance $\sigma^2 = np(1-p)$. Furthermore, normal distributions can be added together to form another (approximate) normal distribution; for example, if we have a normal distribution representing the number of each type of student who will enroll, then we can add together these normals to arrive at a normal distribution representing the entire freshman class.

Therefore, to deal with the uncertainty in the objective function, we might simply take the expected value (mean) of the normal distribution representing the objective measure. For example, if a student of type 1 generates \$4000 in revenue and a student of type 2 generates \$6000 in revenue, and the mean of the normal distributions of types 1 and 2 are each 50 (that is, $np = 50$), then revenues are also normally distributed with mean \$4000(50) + \$6000(50) = \$500,000. This expected value can then be compared with other admissions decisions that yield different means to decide which is most desirable.

The properties of the normal distribution can also be exploited to help us deal with capacity constraints based on the uncertain number of students who enroll. For example, if we wish for the total size of the freshman class, X, to be no larger than 200 with probability 95%, then we can write: $P(X \leq 200) \geq 0.95$. However, if X is normally distributed with mean μ and standard deviation σ, then we can also write: $P(Z \leq (200 - \mu)/\sigma) \geq 0.95$, where Z is a standard normal random variable (see Charnes & Cooper, 1959, 1962, 1963). By employing a normal table, this implies that $(200 - \mu)/\sigma \geq 1.645$. Since μ and σ are functions of the number of students admitted ($n = x_i$), then this constraint can be written in terms of x_i to help determine how many students of a given type should be admitted to ensure that the capacity is not exceeded.

This is the essence of the original model, and much of the essence remains in the refined model to be discussed next. The refinement of the model includes the addition of some new goals and objectives, the reframing of quality constraints in terms of goals the institution wishes to meet, and the relaxation of the assumption that the behavior of students in terms of retention and graduation must be constant from one year to the next. We now discuss the refined model in greater detail.

Refinement of the Original Model

While the original model did a fairly good job of capturing the aspects of the real-world problem of admissions decisions, there are some adjustments that can be made to make the problem more realistic. One of the main changes from the original to the revised model is the reframing of the quality constraints (that the retention and graduation rates must be at least some lower bound with a high degree of probability) as goals to be strived for in the objective function. This decision was made after consulting the chief admissions officer of the participant university, who assured us that it is more realistic to frame these lower bounds as goals, rather

than requirements. Therefore, in the refined model, the only constraints are the real constraints (specifying that no more students can be admitted than applications are present) and the capacity constraints on the number of students enrolled.

Another change from the original model is in the assumption that the behavior of students must not change statistically from year to year. This assumption is still technically needed to ensure validity of the model, since at each point in time we are using current retention and graduation rates to predict behavior of students in the future. However, the original model also proceeded on the assumption that a multi-stage model would eventually be employed, meaning that decisions would not be made solely on a year-by-year basis, but perhaps made while considering several years into the future. It has since been shown that such a complex multi-stage model may not be necessary to adequately describe the problem, and therefore that good decisions may be made when only considering the present year. Therefore, the assumption of the retention and graduation rates not changing from year to year is less vital to the revised model, which now only uses that assumption to try to predict how the current group of students will act in the future. In subsequent years when further decisions are made, the revised model allows for the user to update retention and graduation rates as needed.

In addition to rethinking the quality constraints and relaxing the assumption on the constant enrollment behavior, the main changes in the revised model come in the objective function. Specifically, whereas the original model made mention of various objectives but concentrated on revenues as a measure of desirability of a class of students, the revised model makes explicit many other possible objectives and goals. The revised model concentrates heavily on the quality of a class of students. The quality measures and/or goals in the revised model include:

- Average high school GPA and/or average SAT (or ACT) scores
- Retention and graduation rates and the percentage of students who eventually graduate

- The goal that the freshman class consists of at least or at most a given percentage of students of certain types (for example, no more than a given percentage of "remedial" students or at least a given percentage of high ability students, etc.)
- The goal of enrolling a given number of students of certain types
- The utility associated with a group of students, where utility measures some intangible aspects of a student body.

Average High School GPA or SAT scores

The average high school GPA or SAT scores of an incoming freshman class will depend on the number of students who enroll, which is an uncertain quantity that can be represented by an (approximate) normal random variable. Therefore, if we would like the expected (mean) GPA of the freshman class to be at least some minimum value, then prospective classes that meet this minimum should be rewarded with a positive addition to the objective function, while those that do not meet the minimum should be penalized with a negative addition to the objective function. To establish whether a prospective class meets the minimum, we need to consider the average GPA of the class, which can be estimated by taking a weighted average of the GPAs of the various student types; that is:

$$\text{Average HSGPA} = \overline{G} = \frac{\sum_i G_i X_i}{\sum_i X_i}$$

where G_i is the average high school GPA of students of type i, and X_i is the number of students of type i who enroll. Note that X_i has an approximate normal distribution, so that $\sum G_i X_i$ and $\sum X_i$ are also approximately normally distributed. Then, we would like $\overline{G} \geq G_{\min}$ which means that $\sum G_i X_i - G_{\min} \sum X_i \geq 0$. It can be shown that $\sum G_i X_i$ and $\sum X_i$ have an approximate bivariate normal distribution, and so when added together form another normal

distribution whose mean and variance are based on the means, variances and covariance of the two sums (which are themselves functions of $n = x_i$ and $p = y_i$). Thus, we can simply take the expected value (mean) of the random variable $\sum G_i X_i - G_{min} \sum X_i$ and add it to the objective function. If $\bar{G} \geq G_{min}$, then this will be a positive quantity, otherwise, it will be negative. Note that the idea can be applied to average SAT (or ACT scores).

Retention and Graduation Rates

A similar approach can be taken to measure the expected retention and graduation rates of a class of students, which again depend on the random number of students who enroll and continue to enroll. Let us first consider the first year retention rate of a group of students, which is the percentage of students enrolling in the first year who also enroll in their second year. If we let $X_i(1)$ represent the number of students of type i enrolling in their first year and $X_i(2)$ represent the number enrolling in their second year, then the class's first year retention rate is:

$$\text{First year retention rate} = R^1 = \frac{\sum_i X_i(2)}{\sum_i X_i(1)}.$$

This ratio of normal random variables can be handled in much the same way as average GPA, which was discussed above. Note that the same approach can be taken with second, third, or higher year retention rates, or with any of the graduation rates.

Probability of Graduation

Another possible measure of quality is the percentage of students who are expected to eventually graduate (i.e. the completion rate). The process whereby a student is admitted, enrolls, continues to enroll, and then either graduates or drops out can be described with a mathematical model known as a Markov Chain. Basically, a Markov chain has the property that where the process will go in the next time period will depend only on where the process is in the

current period. In the context of this application, this means that we can track the probability that a student will continue to enroll, will graduate, or will drop out in the next period based only on his status in this period. For example, if a student is enrolled in his first year, then he may enroll in his second year with a probability given by the first year retention rate of students of his type, and may drop out with one minus that probability. If a student is enrolled in his fourth year, then he may graduate after that year with probability equal to the 4-year graduation rate of students of his type, and so on. Basically, if one takes into account each possible year that a student may be enrolled (say years 1 through 7 or 8), and all the retention and graduation rates, a "transition matrix" can be formed that shows the probability that the student progresses to various states in subsequent years. Based on this matrix, one can use Markov Chain methods to determine the probability that the student will eventually graduate.

Knowing the proportion of students of type i who are expected to graduate, let's call it g_i, we can then determine the proportion of students in the whole freshman class who are expected to graduate by taking a weighted average of the g_i's. Letting X_i represent the normally distributed number of students of type i enrolling in their first year, we have:

$$\text{Proportion of students of type i in freshman class} = \frac{X_i}{\sum_i X_i}.$$

So then, the probability of graduation for any student in the class is:

$$P = \sum_i g_i \frac{X_i}{\sum_i X_i} = \frac{\sum_i g_i X_i}{\sum_i X_i}$$

We again have a ratio of normal distributions that can be handled as GPA was handled above.

Bounds on Proportions of Students

Another type of goal that an institution may have is that the freshman class be made up of at least or at most a given percentage of students of a given type. For example, the institution

may want to limit the percentage of the class who are remedial students, have at least a given percentage high ability students, or in the case of public institutions, have at least or at most a given percentage of students who are out-of-state residents. For example, if we let H be the subset of student types representing high ability students, then the percentage of students who are high ability can be represented as:

$$\text{Proportion of freshman class who are high ability} = \frac{\sum_{i \in H} X_i}{\sum_i X_i}$$

where the sum in the denominator is taken over all student types. Again, this ratio can be dealt with in the same way as average GPA above. This reasoning also applies to the proportion of the class who are remedial students, non-residents, or in fact, any subset of student types.

Enrollment Goals

Some institutions may have enrollment goals under which they strive to enroll a given number of students of certain types, for example, an institution may strive to enroll a given number of transfer students. Again, whether or not these goals are met depends on the random number of students who will enroll if admitted. Letting X_i represent the normally distributed number of students of type i enrolling in their first year, we may wish to have that $X_i - b \geq 0$, or that $\sum X_i - B \geq 0$, where the sum could be over all student types or over a subset of student types (for example, all out-of-state residents). If the expected value (mean) of X_i is greater than b, then a positive amount will be added to the objective function, otherwise a negative amount will be added. Since if X_i is normally distributed then so is $\sum X_i$, the same argument holds for $\sum X_i$.

Utility

The final way in which to measure the desirability of student types is to assign a utility value to each type. This utility value can represent any characteristics of the students that are

intangible, are not accounted for in the other desirability measures, or are otherwise hard to quantify. One instance in which an institution may wish to include a generic utility measure is if a group of students helps fulfill an institution's mission (e.g. "access"), but the other measures of quality of the group are such that the group would not be desirable. Another possible example is a group of high ability students for whom the institution may be able to use for marketing purposes, or a group who are expected to contribute to the university in ways other than academics (e.g. sports, campus activities, etc.).

Utility can be measured in various ways, but perhaps an easy way to think about utility is to rate the group's "intangible" assets on a scale from 1 to 10, with 10 being the highest. Then, even students who seem undesirable with respect to quality measures (e.g. students admitted under the "access" mission) will have some kind of positive effect on the objective function, thus encouraging the institution to admit them in some instances.

Once the utility of each student type i is assigned, let's call it u_i, we note that the total utility gained from all students in a given year k is given by $\sum_i u_i X_i(k)$ where $X_i(k)$ is the number of students of type i enrolling in their k^{th} year. Then, the total utility over the group of students' entire tenure at the institution is given by: $\sum_k \sum_i u_i X_i(k)$. Since each $X_i(k)$ is normally distributed, then so is the sum, and so we can simply take the expected value of the sum to arrive at the quantity to be added to the objective function.

Putting it All Together

In each of the objectives and goals above, we were able to take the expected value (mean) of a normal random variable as the contribution to the objective function. Although some of these normal distributions are simpler than others, they are all based on the random variables $X_i(k)$, which represent the number of students of type i enrolling in year k, for all student types i,

and years $k = 1, 2, 3, \ldots$ Each $X_i(k)$ has a mean and variance that is a function of $n = x_i$, and $p = y_i$ or $p = y_i R_i^k$; specifically, the mean of each $X_i(k)$ *is* $np = x_i y_i$ or $x_i y_i R_i^k$. Since all of the normal random variables consist of sums of normals $X_i(k)$, then all of these sums have means of the form $\sum x_i y_i$ or $\sum x_i y_i R_i^k$. In other words, each of the means is a linear function of the decision variables x_i. Therefore, when we multiply each of these objectives by constants and add them together, we arrive at an objective function that is still linear in the decision variables x_i. We will write this combined objective function as $\sum c_i x_i$.

As for the constraints, we must have the constraint that we cannot admit more students of a given type than we have applications from that type; this can be written for each student type as: $0 \le x_i \le A_i$ where A_i is the number of applications from students of type i. Furthermore, recall that the capacity constraints that say $P(X \le U) \ge 0.95$ or $P(X \ge L) \ge 0.95$, where U is some upper bound and L is some lower bound, can be expressed using the properties of the normal as $(U - \mu)/\sigma \ge 1.645$ (which can be rewritten as $\mu + 1.645\sigma - U \le 0$), or $(L - \mu)/\sigma \le -1.645$ (which can be rewritten as $\mu - 1.645\sigma - L \ge 0$). Note again that in each of these constraints, μ and σ^2 are functions, in fact linear functions, of the decision variables x_i. Specifically, since each X in these constraints deals with first year enrollment, we have that $\mu = \sum x_i y_i$ and $\sigma^2 = \sum x_i y_i (1-y_i)$, and so $\sigma = \sqrt{\sum x_i y_i (1-y_i)}$.

Since we wish to maximize the objectives subject to the given constraints, we can write the mathematical program representing this problem in the following form:

Maximize $\sum c_i x_i$

subject to: $\sum x_i y_i + 1.645\sqrt{\sum x_i y_i (1 - y_i)} - U \le 0$

$\sum x_i y_i - 1.645\sqrt{\sum x_i y_i (1 - y_i)} - L \ge 0$

where there may be possibly several capacity constraints, and the sums in those constraints may go over all student types or some subset of student types. In mathematical programming terminology, this problem is called a "non-linear program" since the constraints are non-linear functions of the decision variables x_i. This type of problem can be solved in Excel using an add-in called "Solver." We will discuss how Solver was used to solve specific instances of the problem in the following section.

Applying the Model to Actual Data

Now that all of the revisions to the original model have been discussed in detail, the application of the refined model to actual data can be illustrated. As with the original model, admissions data from a medium-sized, public university in the Midwest were used to establish the student types. In the original model, 25 types were established using an involved analysis of almost 30,000 student records covering the period from 1991-1999 (DePaolo, 2000). After further analysis, 28 types were settled on for the refined model, with two additional types, transfer students and spring freshmen, also included even though historical data for these types were not kept as far back as for the traditional freshmen. Furthermore, the yield, retention and graduation rates were established using this historical data.

Once this preliminary analysis was done, the goals and quality measures for the university were established by consulting the chief admissions officer. This individual identified revenue figures for various student types (in-state vs. out-of-state, on-campus vs. off-campus, etc.), settled on utility values, and established enrollment goals for fall freshmen, transfer students, and the overall undergraduate student body. He also established quality goals pertaining to average high school GPA and SAT scores of the freshman class, and expected first year retention, 4-year graduation, and completion rates for the incoming freshmen. Lastly, he established goals for the

minimum percentage of high ability students, the maximum percentage of so-called "opportunity" students, and a minimum proportion of students from out-of-state.

When questioned about constraints, the admissions official opted for lower and upper bounds on the size of the freshman class, and determined that there could be no more than a given number of students in the university's "opportunity" program due to the resources required to maintain that program. In addition, he identified two additional capacity constraints: one for on-campus freshmen and one for all on-campus undergraduates due to the capacity in the freshmen dormitories and all of the dormitories, respectively.

Once the problem was specified, various incarnations of the problem were solved using applications data from the previous three years, as well as simulated applications data meant to forecast possible future scenarios. The various incarnations involved including or excluding various objectives from the problem. We discuss now the details of these problem solutions.

Solving the Model Using Excel Solver

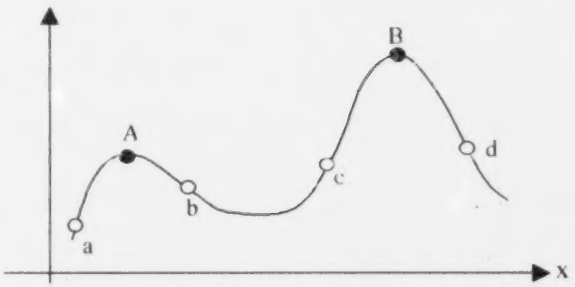
As was mentioned previously, an Excel add-in exists that has the ability to solve non-linear programs. This software was employed to solve several instances of the optimal enrollment problem using data from the participant university. Before discussing the results of those analyses, we give a brief introduction of the Solver software used to conduct the analyses.

In order to use the Solver software, a model of the non-linear program must be set up in an Excel spreadsheet such that cell formulas are used to compute the values of the constraints and the objective function value for a solution consisting of the values of the x_i. In other words, for each possible solution of admittance of a certain number of students of each type, a formula is used to determine if the constraints are met and what the value of the objective function is. Then, Solver attempts to change the values of the x_i in a specific way in order to improve upon the objective function value, until finally, no improvements can be made. When this occurs,

Solver stops and outputs the current value of the x_i's, which represent the number of students of type i that should be admitted.

One drawback to using Solver to solve non-linear programs is that the output may not be the overall optimal solution; that is, just because Solver stops at values of the x_i does not mean that these x_i's cannot be improved upon. It simply means that they cannot be improved upon starting from these particular x_i's. In such a case, the current solution is better than all the points near it, and as such is called a "local optimum." However, what we are really seeking in the solution to this problem is what is called a "global optimum" which is a solution that is better than ALL other possible solutions (see Figure 1). In order to find a global optimum using Solver, one must solve the non-linear program starting from several possible initial solutions containing different values of the x_i. Then, if Solver outputs more than one solution when starting from these different initial solutions, then one can be reasonably (but not completely) sure that the one with the highest objective function value is the global optimum.

Figure 1



A simple one-dimensional illustration shows the difficulty with using Solver on non-linear programs. Depending on the initial point (a, b, c, or d), Solver may converge to the local optimum A and stop, thinking no improvements can be made, when B is in fact the best solution.

Applying the Model to Different Admissions Strategies

This approach to solve this problem with Solver was taken for a total of 267 instances of the problem using the university's actual applications data from previous years, and another 105 problems using simulated applications data. When simulating applications, it was assumed that applications of each student type were normally distributed. The means and standard deviations for these distributions were estimated from actual admissions data from the past several years. The problems fell into several categories, and were representative of several different strategies that an institution might employ. We discuss some of those strategies now.

1. Revenues Only: One strategy that an institution might employ is to try to maximize revenues. This type of strategy was implemented in the Excel model by eliminating quality and enrollment considerations from the objective function. The optimal admissions policy in this case is exactly what one would expect: admit as many students as possible, except when one of the upper bounds on capacity is violated. If this upper bound is violated, then students from the least desirable student types would not admitted. We will discuss later how the "least desirable" student type is identified.
2. Enrollment Goals and Utility Only: Strategies that seek to meet enrollment goals and maximize utility (where every student lends some positive utility to the objective function) turned out to have optimal solutions similar to those maximizing revenue only. Specifically, to meet enrollment goals or maximize utility, the optimal admissions strategy is to admit as many students as possible from all types, unless an upper bound on capacity is violated, in which case only the least desirable student types should be denied admittance. The only difference between these strategies and the revenue-based strategy is in how the "least desirable" student types are determined, since here we want least desirable with respect to enrollment/utility as opposed to least desirable with respect to revenues.
3. Quality Only: Another possible strategy that may be employed by an institution is to maximize quality while ignoring other considerations such as revenue and enrollment. This strategy can be implemented in many different ways, depending on which measures of quality are included or excluded from the objective function. The optimal solutions to these problems were also very intuitive: they suggested that only the highest quality students should be admitted, unless a lower bound on capacity is violated (that is, having only the best students leaves the freshman class unacceptably small). In this case, the "best" students who are remaining should be admitted, but only until the class is acceptably large. Again, which

groups are determined to be the "best" students depends on which measure of quality is used, since some groups may, for instance, have high SAT scores, but low retention rates.

4. Combinations of Revenues and Quality: Policies concentrating on revenue result in large numbers of students and correspondingly low measures of quality, while policies emphasizing quality result in very small classes that may not generate enough revenue for the institution to function effectively. One way to incorporate both ideas is to include both objectives in the model and give each a certain amount of weight. Several such combination strategies were employed. Some weighted revenues highly, some weighted quality highly, and some combined the two with equal weight. The results were as would be expected – by varying the weight given to each objective, solutions could be made to give as large or as small classes with as high quality standards as desired. These solutions could be useful to an institution to try to determine what the "cost" would be in terms of enrollment if they wished to increase quality by a certain amount.

Properties of the Optimal Solutions

It is interesting to note that the solutions to these 372 problems had some common characteristics. For instance, only two of the five constraints specified by the admissions official had any effect on the optimal solution. The three constraints that had no effect on the solutions to the problem were the ones pertaining to the capacity in the dormitories, and the one placing an upper bound on the size of the freshman class. Thus, it can be concluded that if the applications patterns over the last several years continue in the future, these constraints are not likely to be an issue to be considered when making admissions decisions.

However, the two constraints that were of significance were the lower bound on the size of the freshman class, and the upper bound on the number of "opportunity" students that the institution can effectively serve. As might be expected, the constraint for the lower bound on the

size of the freshman class tended to be "active" when quality measures weighed heavily in the objective function, whereas the constraint on the number opportunity students came into play when revenues, enrollment, or utility were the focus of the admissions strategy. (An active inequality constraint is one that is met with equality at the optimal solution.) In a few cases, both of these constraints were important in the solution to the problem.

Another property that all of the solutions had in common is that they were "greedy" in nature. In mathematical programming terms, a "greedy" solution gets its name from the fact that the solution has an "all or nothing" structure. In this case, the "greedy" solutions to this non-linear program were such that almost all of the values of the x_i's in each optimal solution were either at their lower bound, which is zero, or their upper bound A_i, which is the number of applications from students of that type. If only one of the above-mentioned constraints came into play in the optimal solution, then only one of the x_i's was between these bounds, and if two of the constraints were active, then two x_i's were between the bounds.

The structure of these solutions can be explained in an intuitive way. Suppose quality is the only objective being considered. Then you would want only the highest quality students to be admitted, but if that resulted in an unacceptably small freshman class, then the best of the remaining students would need to be admitted. However, since you really only want these "filler" students to get to your lower bound, you would not want to admit any more of them than is necessary to achieve that goal; therefore, you would only admit enough of them to equal your lower bound. This may mean that some students of a given type are admitted, while the rest are not. The same reasoning would apply when an upper bound comes into play when, for example, revenues are the focus of the strategy. You want as many students as possible, but you cannot violate an upper bound, so you may have to stop when some students of a given type are admitted and the rest are not.

That these solutions all shared this property is not simply a matter of intuition. It can be proved mathematically that the structure of the problem always leads to solutions with such a "greedy" structure. However, the presence of this structure means that knowing which students are the "best" or "worst" is important to finding the best solution; that is, in order to find where to "stop" admitting students, we need to establish an ordering that will tell us which student types are most and least desirable. Finding this ordering is the topic of the following section.

Ordering the Student Types

In order to establish an "ordering" of student types, let us first consider the case when only one of the constraints will be active at the optimal solution, say the upper bound on some subset of student types, which we will denote by S. Which types will be most desirable in this case? Let us break this down in a cost/benefit framework, or in other words, let us find which students give us the most "bang for the buck."

Obviously, we would want to admit students who contribute positively to the objective function, that is, types i for which $c_i > 0$ (the "bang"). Furthermore, student types that are not in the subset S are more desirable than those who are since they "cost" us nothing. Therefore, we begin our "best" to "worst" ordering by considering only those students who are not in the subset S and for whom $c_i > 0$, and order them from largest c_i to smallest. Note that student types that are not in S and for which $c_i \leq 0$ are not desirable since they detract from our objectives, so we will put these students at the end of the list, with the smallest (most negative c_i) at the end.

As for the student types that are in S, they may contribute to our objectives, but that contribution comes at a price – namely, how much they contribute to the constraint. For example, there are many student types that we may want to admit because they generate revenue (the "benefit") but they also may "cost" us in terms of capacity. Therefore, what we really want

to do for those student types in S is to order them (in the middle of our list) from the highest to lowest benefit/cost ratio, or in other words, from largest to smallest "bang for your buck."

The benefit portion of the benefit/cost ratio is not difficult to discern – it is the value of c_i, the contribution of one student of type i to the objective. However, it can also be thought of as the "rate of change" of the objective function with respect to the decision variable x_i, which is also the derivative of $\sum c_i x_i$ with respect to x_i. In a similar manner, we can look at the rate of change of the "cost" with respect to x_i, which is the derivative of the constraint with respect to x_i. Thus, recalling that we are dealing with an upper bound, we have:

$$\frac{\partial}{\partial x_i}\left[\sum x_i y_i + 1.645\left[\sum x_i y_i (1-y_i)\right]^{1/2} - U\right] = y_i + 1.645 \cdot \frac{1}{2} \cdot \left[\sum x_i y_i (1-y_i)\right]^{-1/2} \cdot y_i (1-y_i) = y_i + \frac{1.645 y_i (1-y_i)}{2\sqrt{\sum x_i y_i (1-y_i)}}$$

Therefore, the correct benefit/cost ratio is: $\dfrac{c_i}{y_i + \dfrac{1.645 y_i (1-y_i)}{2\sqrt{\sum x_i y_i (1-y_i)}}}$.

These ratios for student types in S should be ordered from highest to lowest and placed in between the positive c_i's and the negative c_i's on the list started above.

Note that a similar approach can be taken when the constraint of interest is the lower bound, for example, when quality is of the utmost concern in the objective function. However, the approach is slightly different because here we *want* students who "cost" us something in the constraints because we are seeking to meet a *lower* bound. Thus, the first thing to do is to admit all of those students who have a positive effect on the objective function, that is, the order begins with student types for which $c_i > 0$ being ordered from largest c_i to smallest. Then, if we admit all of these students and the lower bound is not met, we must then admit students who have a zero or negative effect on the objective function, starting with those with a zero effect. Following this reasoning, next on the list are those students for whom $c_i = 0$ and who will contribute to the lower bound (a good thing) but *not* the upper bound (a bad thing), so we will

order these students in terms of the derivative of the lower bound constraint with respect to x_i; that is, from the highest to lowest values of:

$$\frac{\partial}{\partial x_i}\left[\sum x_i y_i - 1.645\left[\sum x_i y_i (1-y_i)\right]^{1/2} - L\right] = y_i - \frac{1.645 y_i (1-y_i)}{2\sqrt{\sum x_i y_i (1-y_i)}}$$

The next most desirable students are those for which $c_i = 0$ and which contribute to the lower *and* the upper bound. Here, we wish to take into account the contribution to the lower bound, which we want to be large, and the contribution to the upper bound, which we want to be small. Therefore, the next step is to order these types from the highest to lowest values of:

$$\frac{y_i - \dfrac{1.645 y_i (1-y_i)}{2\sqrt{\sum x_i y_i (1-y_i)}}}{y_i + \dfrac{1.645 y_i (1-y_i)}{2\sqrt{\sum x_i y_i (1-y_i)}}}.$$

Then, if the lower bound is still not met, we must unfortunately admit some students with a negative effect on our objective simply because we must meet our lower bound. For those student types for which $c_i < 0$, we again consider the benefit/cost ratio by ordering these from largest to smallest:

$$\frac{c_i}{y_i - \dfrac{1.645 y_i (1-y_i)}{2\sqrt{\sum x_i y_i (1-y_i)}}}.$$

(Recall that here $c_i < 0$, so the largest ratio is actually the least negative.)

Solving the Problem Using Type Orderings

Once we have these orderings of the student types, and armed with the knowledge that the solution must be greedy in nature, it is a fairly straightforward process to estimate the optimal solution. We simply go down our list, admitting all of each subsequent student type, until one of the constraints is violated. We then "back up" so to speak, determining exactly where between 0 and its upper bound x_i needs to be in order for that constraint to be met with equality. This idea forms the basis for an algorithm to solve this type of problem *without using Excel Solver*.

However, there is one problem with this approach. If we inspect the ratios described above that are necessary to establish the orderings, we note that they contain values of x_i, which are actually the values of the x_i *at the optimal solution*; however, we can't know the optimal solution unless we establish the orderings! This is quite a dilemma, and one that actually makes it impossible to know with certainty the correct ordering before the problem is solved. Nonetheless, it is possible to estimate the values of the x_i's so that we can establish an *approximate* ordering that will be very close to the correct order. We can then use this approximate order to implement our algorithm and arrive at a solution that may not be the global optimum of the problem, but which has been shown to be extremely close to that optimum.

The basis of this approximation approach is as follows: if we know that either the lower or upper bound will be met at equality at the optimal solution, then the value of $\sum x_i y_i$ will be very close to that bound. This approximation works because our constraint is essentially stating that the mean of a normal distribution plus or minus a certain number of standard deviations is equal to a bound. Thus, assuming that the standard deviation is relatively small, this implies that the mean, $\sum x_i y_i$, is roughly equal to that bound. Furthermore, we can estimate the terms $(1 - y_i)$ with the term $(1 - \bar{y})$ where $\bar{y}$ is the average of those values of y_i to get:

$$\sum x_i y_i (1 - y_i) \approx (1 - \bar{y}) \sum x_i y_i \approx (1 - \bar{y}) B$$

where B is either the value of the upper bound U or the lower bound L. Whether or not B is equal to U or L can be guessed by observing whether quality or revenues/enrollment are the dominating objective; if quality dominates, the lower bound will likely be active, whereas strong emphasis on revenues and enrollment will likely cause the upper bound to be active. This guess can then be confirmed by setting all of the x_i's for which $c_i > 0$ to their upper bound and

observing whether the lower or the upper bound is violated. The value of B should be set equal to the value of the bound that is violated.

Again, this approximation will allow us to establish an ordering that is very close to, but not necessarily equal to the exact ordering. Despite the possibility that it may not establish the correct ordering, this procedure is very useful in approximating the solution to the non-linear program. As we will discuss in the next section, this so called "approximation algorithm" for solving these problems is simple to implement in Excel and requires no add-in software. Furthermore, the algorithm arrives at the optimal solution in almost all of the cases, and in those where it does not, gives a solution that is almost imperceptibly inferior to the actual solution.

Implementing the Algorithm in Excel

The aforementioned approximation algorithm can be implemented in Excel by first setting up the spreadsheet to account for the parameters of the model, including the objective coefficients and the yield, retention and graduation rates of each student type. Then, one can simply specify the objective function to be utilized and the number of applications of each type under consideration, and the algorithm will immediately output the solution.

This approximation was tested on the 372 problems previously solved by Excel Solver, and the solutions were compared. In the 267 problems based on actual application data from past years, the algorithm found the optimal solution established by Solver in 262 of the cases. In the remaining 5 cases, the algorithm found a solution that was extremely close to the optimal (over 99.9997% of the optimal objective function value). This result implies that, although the algorithm did not find the global optimum, it found a solution that, for all intents and purposes, is equivalent to the optimal solution. In general, these sub-optimal solutions were found because two or more student types were extremely close to one another in their ratios, but because the

approximation algorithm approximated the denominator of those ratios, the ordering became reversed in the approximate ordering. In other words, in the optimal solution, one type was taken before the other, while in the approximation, the second was taken before the first. Although this results in two different solutions to the problem, we see that the objective function values are almost equal because the two types are almost indistinguishable.

In the remaining 105 problems solved using simulated numbers of applications, the approximation algorithm found the optimal solution in all cases. This difference could be due to the fact that simulated problems are usually not to be as extreme as some actual problems might be, and it is usually these extreme problems in which the algorithm falters.

From this analysis, it seems safe to conclude that the approximation algorithm works very well in a vast majority of cases. In addition, it is very fast. Solving a problem using the algorithm takes an average of 1/3 of a second on a 500 MHz Pentium III PC with 128 MB RAM. Solving the same problems using Excel Solver takes an average of 3 to 4 seconds on the same machine *for each initial solution examined.* (Recall that several initial solutions should be investigated to be sure that Solver has stopped at the global optimum.) Therefore, the algorithm is at least 10 times faster than using Solver.

Using the Model for Forecasting

While the main application of the algorithm is to determine the best course of action when faced with a pool of applicants for which admissions decisions must be made, the model also has the added benefit of being easily adapted to provide forecasts. For example, an institution might wish to see what would happen to various measures of revenue, enrollment and quality if an admissions strategy were to be implemented for several years. To do this, the institution would only need to use Excel to simulate applications for those years and use the algorithm to project what the admissions decisions would be under that strategy in each year. Then, quantities like

revenues, enrollment figures, retention and graduation rates, and average GPA/SAT scores could be measured in the simulated scenarios to give an idea of where that strategy would lead in the long run. Furthermore, several different strategies could easily be implemented, the measures calculated, and the outputs compared. Due to the speed of the algorithm, this entire process could be done in a few hours on the PC described above.

This process was implemented for the participant university's data, using 13 different strategies including revenue only, quality only, enrollment and utility only, and various combinations of these measures. For each strategy, 100 different simulations were run, each randomly generating 10 years worth of applications data according to a normal distribution, and the algorithm was used to solve the problem in each year. Then, summary measures for each scenario were computed by taking the average over the 100 simulations. These measures included: revenues, total enrollment, freshman enrollment, average GPA/SAT of the freshman class, retention and graduation rates, percentage of freshman exceeding minimum admissions requirements, and percentage of the freshman class who are high ability, enrolled in the opportunity program, and out-of-state residents.

When one compares the simulation results from strategies that maximize revenues and enrollment to those that emphasize quality only, some startling results emerge. For the data used, maximizing quality would, in the long run, result in enrollment and revenues dropping roughly 50% compared to cases in which enrollment and revenues are maximized. Obviously, a strategy such as this would not be fiscally viable for most institutions.

However, as would be expected, this loss of revenues and enrollment led to increased quality. Average GPA of the freshman class increased by roughly 0.16, average total SAT scores increased by about 50 points, first year retention rates increased by about 4%, and graduation rates increased by about 5%, depending on the specific strategy. In some respects,

these increases do not seem drastic; however, we must recall that all these measures assume that the applications received in the future will follow the same patterns as have been observed in the past. Thus, as long as the applicant pool does not change significantly, it will be very difficult to drastically alter quality measures.

While only concentrating on revenues and enrollment deteriorates quality, and only focusing on quality appears to have catastrophic financial ramifications, there is much to be learned from other scenarios that combine revenues, enrollment, and quality in various ways. The four strategies tested that combine all of these objectives seem to offer reasonable compromises between these two extremes from which an institution can estimate the trade-off between revenues and quality. For example, under one strategy, the institution might forfeit a small percentage of its profits to increase quality in several areas, while in another, a larger drop in revenues might mean significant gains in certain key areas. Armed with this information, an institution can begin to comprehend the long run consequences of a change in admissions strategy, and make informed decisions as to whether such a change is in its best interests.

Conclusions

This mathematical model is a useful formulation of the optimal enrollment problem because it takes into account many issues that an institution may want to consider when making admissions decisions, including revenues, enrollment, and various quality measures, while considering capacity constraints that are imposed on the institution. The model integrates the uncertainty associated with the enrollment, persistence and graduation of students, and allows it to be dealt with effectively. Furthermore, the formulation allows the problem to be solved with established mathematical programming methods.

The application of the model to actual application data from the participant university as well as simulated application data meant to approximate future applications shows that the model

performs well and provides reasonable solutions that make intuitive sense to the user. Furthermore, the solution of this model using Excel Solver points to a special structure that all solutions to the model share: specifically, that each optimal solution is "greedy" in nature, meaning that the optimal solution will be such that for all except one or two student types, it is best to admit either all or none of these students. This information in and of itself may be useful for admissions officials since it determines which students are good and bad risks.

Furthermore, the framework of this model allows us to exploit this special structure of the solutions in special cases, namely when only one or two bounds are active at the optimal solution. In this case, it is possible to establish an "ordering" of the student types, from best to worst, whereby the optimal solution involves simply going down the list, and admitting all of the students of each subsequent type, until the bound is met with equality. This ordering provides a basis on which to base an approximation algorithm to solve the problem. This algorithm finds the optimal solution, or extremely close to it, in essentially all of the cases, and has the added benefit of being fast and easily implemented in Excel without any add-ins or knowledge of mathematical programming. Furthermore, this algorithm, combined with simulation techniques, can be used to look several years in the future to see what the long-term effects of an admissions strategy may be.

Together, the model and its solution algorithm appear to be a tool that would be useful to any institution that struggles with the problem of optimal enrollments. The model takes into account many different aspects of the problem and expresses them in a meaningful way, while the algorithm provides insight into the results in terms of orderings of student types and possible long-term effects of admissions strategies. The ability to use Excel, software to which most institutions have access, to implement the model quickly and easily and to run many different scenarios to compare results give the model an added benefit.

References

Averill, R.F., & Suttle, J.L. (1975). A Model for University Enrollment Planning. Socio-Economic Planning Sciences, 9, 257-261.

Bazaraa, M., Sherali, H., & Shetty, C. (1993). Nonlinear Programming: Theory and Algorithms. New York: John Wiley & Sons.

Brinkman, P.T., & McIntyre, C. (1997) Methods and Techniques of Enrollment Forecasting. In D.T. Layzell (Ed.), Forecasting and Managing Enrollment and Revenue: An Overview of Current Trends, Issues and Methods. New Directions in Institutional Research 93, (pp. 67-80). San Francisco: Jossey-Bass Publishers.

Bruggink, T.H., & Gambhir, V. (1996). Statistical Models for College Admission and Enrollment: A Case Study for a Selective Liberal Arts College. Research in Higher Education, 37, (April), 221-240.

Caruthers, J.K., & Wentworth, C.L. (1997). Methods and Techniques of Revenue Forecasting. In D.T. Layzell (Ed.), Forecasting and Managing Enrollment and Revenue: An Overview of Current Trends, Issues and Methods. New Directions in Institutional Research 93, (pp. 81-93). San Francisco: Jossey-Bass Publishers.

Charnes, A., & Cooper, W.W. (1959). Chance-Constrained Programming. Management Science, 6 (1), 73-79.

Charnes, A., & Cooper, W.W. (1962). Chance Constraints and Normal Deviates. Journal of the American Statistical Association, 57, 134-148.

Charnes, A., & Cooper, W.W. (1963). Deterministic Equivalents for Optimizing and Satisficing Under Chance Constraints. Operations Research, 11, 18-39.

Conner, A. (1971). Community College Enrollment Projections. Washington, D.C.: American Association of Junior Colleges.

Day, J.H. (1997). Enrollment Forecasting and Revenue Implications for Private Colleges and Universities. In D.T. Layzell (Ed.), Forecasting and Managing Enrollment and Revenue: An Overview of Current Trends, Issues and Methods. New Directions in Institutional Research 93, (pp. 51-65). San Francisco: Jossey-Bass Publishers.

DePaolo, C.A. (2000). Optimal Enrollments: A Mathematical Approach. Paper presented at the 40th Annual AIR Forum, Cincinnati, Ohio, May 21-24. On the World Wide Web: http://airweb.org/members/Forum00Papers/108.pdf

Donhardt, G.L. (1995). Tracking Student Enrollments Using the Markov Chain, Comprehensive Tool for Enrollment Management. Journal of College Student Development, 36, (September/October), 457-462.

Hossler, D., Bean, J.P., & Associates (1990). The Strategic Management of College Enrollments. San Francisco: Jossey-Bass Publishers.

Ihlanfeldt, W. (1980). Achieving Optimal Enrollments and Tuition Revenues. San Francisco: Jossey-Bass Publishers.

Ingersoll, R.J. (1988). The Enrollment Problem: Proven Management Techniques. American Council on Education/Macmillan Publishing Company.

Kemerer, F.R., Baldridge, J.V., & Green, K.C. (1982) Strategies for Effective Enrollment Management. Washington, D.C.: American Association of State Colleges and Universities.

Schellenberg, S.J., & Stephens, C.E. (1987). Enrollment Projection: Variations on a Theme. Paper presented at the Annual Meeting of the American Educational Research Association, Washington, D.C., April 20-24.